

## Article

# Traffic Volatility Forecasting Using an Omnibus Family GARCH Modeling Framework

Jishun Ou <sup>1,2</sup>, Xiangmei Huang <sup>1</sup>, Yang Zhou <sup>3</sup>, Zhigang Zhou <sup>1</sup> and Qinghui Nie <sup>1,\*</sup><sup>1</sup> College of Architectural Science and Engineering, Yangzhou University, Yangzhou 225127, China<sup>2</sup> State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China<sup>3</sup> Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77840, USA

\* Correspondence: nieqinghui@yzu.edu.cn

**Abstract:** Traffic volatility modeling has been highly valued in recent years because of its advantages in describing the uncertainty of traffic flow during the short-term forecasting process. A few generalized autoregressive conditional heteroscedastic (GARCH) models have been developed to capture and hence forecast the volatility of traffic flow. Although these models have been confirmed to be capable of producing more reliable forecasts than traditional point forecasting models, the more or less imposed restrictions on parameter estimations may make the asymmetric property of traffic volatility be not or insufficiently considered. Furthermore, the performance of the models has not been fully evaluated and compared in the traffic forecasting context, rendering the choice of the models dilemmatic for traffic volatility modeling. In this study, an omnibus traffic volatility forecasting framework is proposed, where various traffic volatility models with symmetric and asymmetric properties can be developed in a unifying way by fixing or flexibly estimating three key parameters, namely the Box-Cox transformation coefficient  $\lambda$ , the shift factor  $b$ , and the rotation factor  $c$ . Extensive traffic speed datasets collected from urban roads of Kunshan city, China, and from freeway segments of the San Diego Region, USA, were used to evaluate the proposed framework and develop traffic volatility forecasting models in a number of case studies. The models include the standard GARCH, the threshold GARCH (TGARCH), the nonlinear ARCH (NGARCH), the nonlinear-asymmetric GARCH (NAGARCH), the GJGARCH (GJR-GARCH), and the family GARCH (FGARCH). The mean forecasting performance of the models was measured with mean absolute error (MAE) and mean absolute percentage error (MAPE), while the volatility forecasting performance of the models was measured with volatility mean absolute error (VMAE), directional accuracy (DA), kickoff percentage (KP), and average confidence length (ACL). Experimental results demonstrate the effectiveness and flexibility of the proposed framework and provide insights into how to develop and select proper traffic volatility forecasting models in different situations.

**Citation:** Ou, J.; Huang, X.; Zhou, Y.; Zhou, Z.; Nie, Q. Traffic Volatility Forecasting Using an Omnibus Family GARCH Modeling Framework. *Entropy* **2022**, *24*, 1392. <https://doi.org/10.3390/e24101392>

Academic Editor: Sotiris Kotsiantis

Received: 21 August 2022

Accepted: 26 September 2022

Published: 29 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Keywords:** traffic volatility; asymmetric property; omnibus family GARCH model; short-term traffic flow forecasting; traffic reliability



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Short-term travel flow forecasting serves as an essential tool to provide useful traffic flow information in the immediate future, and hence wide supports to intelligent transportation systems (ITSs), especially advanced traveler information system, route guidance system, and proactive traffic signal control system. For instance, with accurate real-time traffic flow forecasts, a proactive signal control system is able to set up strategies to significantly reduce traffic delay, and a route guidance system is capable of providing travelers with the optimal route with a minimum travel cost. With this in mind, plenty of

methods have been proposed for short-term traffic flow forecasting, which can be generally divided into two categories: (1) traditional forecasting methods that concentrate on improving the accuracy of traffic mean forecasts; and (2) traffic volatility forecasting that focus on the unpredictable part to capture the time-dependent variances.

During the last couple of decades, extensive traffic flow forecasting studies have been conducted using time series models [1–4], Kalman filtering [5,6], nonparametric regression [7,8], neural networks [9,10], Bayesian network [11,12], random forests [13,14], deep neural networks [15–17], and so on. The majority of the methods belong to point forecasting methods, in which the second moment (i.e., variance) of the traffic variables was presumed as a constant. Therefore, traffic volatility was artificially omitted during the forecasting process.

As the traffic reliability or uncertainty problem has been highly valued in recent years, many volatility models have been proposed for short-term traffic flow forecasting. Compared with the point forecasting models that presume a constant variance, the volatility models can not only produce accurate mean forecasts but also give more reliable forecasting confidence intervals by capturing the temporal evolution of conditional variances. Among the volatility models, the GARCH-type models, such as the fractionally integrated asymmetric power ARCH (FIAPARCH) model [18], the exponential GARCH (EGARCH) model [19], and the GJR-GARCH model [20], have enjoyed much popularity in past years, for their outstanding ability to describe the asymmetric effect of positive and negative shocks on volatility. Even though these models have been confirmed to be able to produce more reliable forecasts than the traditional point forecasting models, the more or less imposed restrictions on parameter estimations may make the asymmetric property of traffic volatility be not or insufficiently considered. Furthermore, the performance of the models has not been fully evaluated and compared in the traffic forecasting context, rendering the choice of the models dilemmatic for traffic volatility modeling. Therefore, it is necessary to develop a flexible traffic volatility forecasting framework under which various volatility models can be easily implemented and fully tested to yield more reliable forecasts and enhance the forecasting models' applicability.

In this research, an omnibus traffic volatility forecasting framework based on the FGARCH model [21] is proposed to model traffic speed volatility on urban roads and freeways. The volatility models under the proposed framework are established by fixing or flexibly estimating three key parameters, namely the Box-Cox transformation coefficient  $\lambda$ , the shift factor  $b$ , and the rotation factor  $c$ . Extensive case studies show that the proposed framework can be used to account for both shift and rotation asymmetric effects and allows for more flexible traffic volatility descriptions that are usually restricted in some classical volatility models. The achieved findings provide insights into how to develop and select proper traffic volatility forecasting models in different situations and are helpful for building traffic flow forecasting models with high reliability and good applicability.

The main contributions of this research include two aspects. First, we present a comprehensive traffic volatility forecasting framework, which can be utilized to develop effective volatility models in a flexible manner, and are capable of properly capturing the evolution of conditional variance with symmetric or asymmetric properties. Second, we conduct extensive experimental studies to fully evaluate and compare the performance of the established volatility forecasting models, including the symmetric standard GARCH [22] and the asymmetric TGARCH [23], NGARCH [24], NAGARCH [25], GJR-GARCH [19], and FGARCH [21] and further summarize their applicability in various situations.

The remainder of this paper is organized as follows: In Section 2, a comprehensive literature review regarding traffic volatility forecasting is provided. Then, the proposed framework based on the FGARCH model for traffic volatility forecasting is elaborated in Section 3. In Section 4, the performance of the developed traffic volatility forecasting models is fully evaluated and compared using real traffic speed datasets collected from urban

road links and freeway segments. Conclusions and future work are summed up in Section 5.

## 2. Literature Review

During the past decades, a large number of volatility methods have been developed for short-term traffic flow forecasting. In general, these methods describe the traffic volatility during the forecasting process by using the standard GARCH or its advanced extensions.

The GARCH-type models are firstly proposed in the fields of finance and economics. In order to deal with the problem that the volatility of financial time series changes over time, the autoregressive conditional heteroscedastic (ARCH) model was introduced by Engle [26], and then further promoted by Bollerslev [22] as the GARCH model. In a following study, Kamarianakis et al. [27] pointed out that the time-varying nature of traffic volatility can be recognized as the alternative large movement period that can be observed when the traffic condition changes. Therefore, a standard GARCH model was applied to study the volatility of relative speed. Given the strong ability to capture the traffic volatility with time-variant conditional variances, the GARCH model has been widely applied in many pieces of research for short-term traffic flow forecasting. Guo and Williams [28] proposed to use the autoregressive moving average (ARMA) model for traffic speed mean forecasting and the GARCH model for traffic speed volatility forecasting. At the stage of online processing, a layered adaptive Kalman filter approach was developed. Similarly, Chen et al. [29] proposed an integrated model which combines the autoregressive integrated moving average (ARIMA) model and the GARCH model for traffic flow forecasting and reached the conclusion that the time-variant confidence intervals achieved can provide more reliable information for travelers, compared with the standard ARIMA model. Ding et al. [30] modeled the dynamic volatility of the subway short-term ridership forecasting process by constructing four kinds of integrated ARIMA-GARCH models. Their experimental results show the performance of the hybrid models is better than the traditional models because of the improvement in the reliability of the forecasting point value with the associated coverage probability of the forecasting interval. Considering that the inherent relationships between different traffic variables could contribute to the improvement of traffic forecasting accuracy and reliability, Xia et al. [31] developed a multivariate GARCH model for traffic volume and speed volatility forecasting, in which a vector autoregressive model was used for jointly forecasting traffic volume and speed mean values. In addition, from the perspective of traffic flow decomposition, Chen et al. [32] proposed a time-series analysis and supervised-learning (TSA-SL) model for short-term traffic flow prediction by decomposing the traffic flow series into two components, namely periodicity and volatility.

Due to the standard GARCH process having a linear relationship between the conditional variance and the shocks in the past, the GARCH model is limited in perfectly describing the temporal evolution of traffic volatility. Moreover, the squared term in the standard GARCH model does not allow the asymmetric effects of traffic volatility that might exist when traffic condition changes dramatically. Therefore, Tsekeris and Stathopoulos [18] presented a fractionally integrated asymmetric power ARCH (FIAPARCH) model to relax the linear and symmetric restrictions of the standard GARCH model, and in return, to better represent the nonlinear and asymmetric properties of traffic volatility. In their research, the fractionally integrated component was introduced to describe the long memory of traffic volatility since the standard GARCH process can only capture the short memory of the volatility evolution that was restrained by the assumption of exponentially decayed propagation of shocks. Evaluation results show that the FIAPARCH model outperforms the standard GARCH model in traffic volatility forecasting. Similarly, Zhang et al. [19] applied another two asymmetric volatility models, i.e., an EGARCH model and a GJR-GARCH model, for reliable short-term travel time forecasting. Compared with the standard GARCH model, both the EGARCH and GJR-GARCH models

allow the conditional variance to respond differently to the past negative and positive shocks that release the symmetric restriction in the GARCH model. Performance evaluation demonstrates that the GJR-GARCH is better than the standard GARCH model and the EGARCH model in terms of the forecasting bandwidth and coverage probability.

Considering traffic volatility may be better modeled by removing underlying traffic patterns that can be decomposed from original traffic flow data series, Zhang et al. [20] proposed a hybrid traffic flow forecasting model that used spectral analysis to remove the periodic trend of traffic flows using an ARIMA process to model the deterministic component of traffic flow, and using a GJR-GARCH process to model the traffic volatility. Experimental results show that the hybrid model is able to improve forecasting accuracy and reliability. Years later, Lin et al. [33] proposed a hybrid PSO-ELM model to quantify the uncertainty of border crossing traffic volume prediction and compared it with the adaptive Kalman filter method used by Guo et al. [28] and the GJR-GARCH method proposed by Zhang et al. [20]. The results confirm that the proposed PSO-ELM model can always reach satisfactory forecasting reliability. As a nonlinear analytical model, the integer-valued GARCH (INGARCH) can be utilized to capture the characteristics of network traffic. Therefore, it was introduced in the modeling of network traffic forecasting by Kim [34] and turned out to be more competitive than the ARIMA, GARCH, and long short-term memory neural network. In addition, Yao et al. [35] presented a nonlinear hybrid method for traffic flow forecasting, which uses the ARIMA-GARCH model to predict the similar and volatile parts, and applies the Markov model with state membership degree and wavelet neural network to predict the irregular part. Another hybrid model named ARIMA-GARCH-M developed by Lin and Huang [36] is able to predict short-term high-speed traffic flow data with good accuracy and reliability. To extend the GARCH model to be used for spatial data, Sato and Matsuda [37] proposed a spatial GARCH (S-GARCH) model. Considering the effect of both space and time restrictions, Hølleland and Karlsen [38] introduced a stationary spatiotemporal GARCH. To deal with the time-varying characteristic of traffic flow, a hybrid method that combines the ARIMA, maximum correntropy criterion, conditional kernel density estimation, and GARCH model was put forward by Zhao et al. [39], which outperforms the ARIMA model in terms of both accuracy and reliability.

The aforementioned GARCH-type models describe the evolution of traffic volatility from linear to nonlinear, from symmetric effect to asymmetric effect. Though the dynamics of traffic volatility has been modeled explicitly step-by-step, the ability of capturing traffic volatility may be subject to the restrictions of the appointed GARCH-type models. For instance, the asymmetric property of volatility can be captured by EGARCH, GJR-GARCH, and asymmetric power ARCH models because of the introduced “rotation factor” in the ARCH term. However, another critical factor—“shift factor”—that can also be used to describe the asymmetric property is not considered in these models. According to Helbing et al. [40,41], the volatile behavior dynamics of drivers was found to be similar to the stockholders in a financial market. As both the rotation and shift effects have been recognized in the financial market [21], it is reasonable and necessary to develop a more flexible volatility model that permits both shift and rotation asymmetric properties for traffic volatility modeling. Moreover, according to Hentschel [21], multiple volatility models are able to be established within the same theoretical GARCH framework. However, the performance of these models on traffic volatility has not been comprehensively tested, making their feasibility in traffic reliability inexplicable for researchers and practitioners.

### 3. Method

To develop an effective traffic volatility model that can better capture the evolution of the conditional heteroskedasticity underlying traffic flow time series, an omnibus traffic volatility forecasting framework is presented, as depicted in Figure 1.

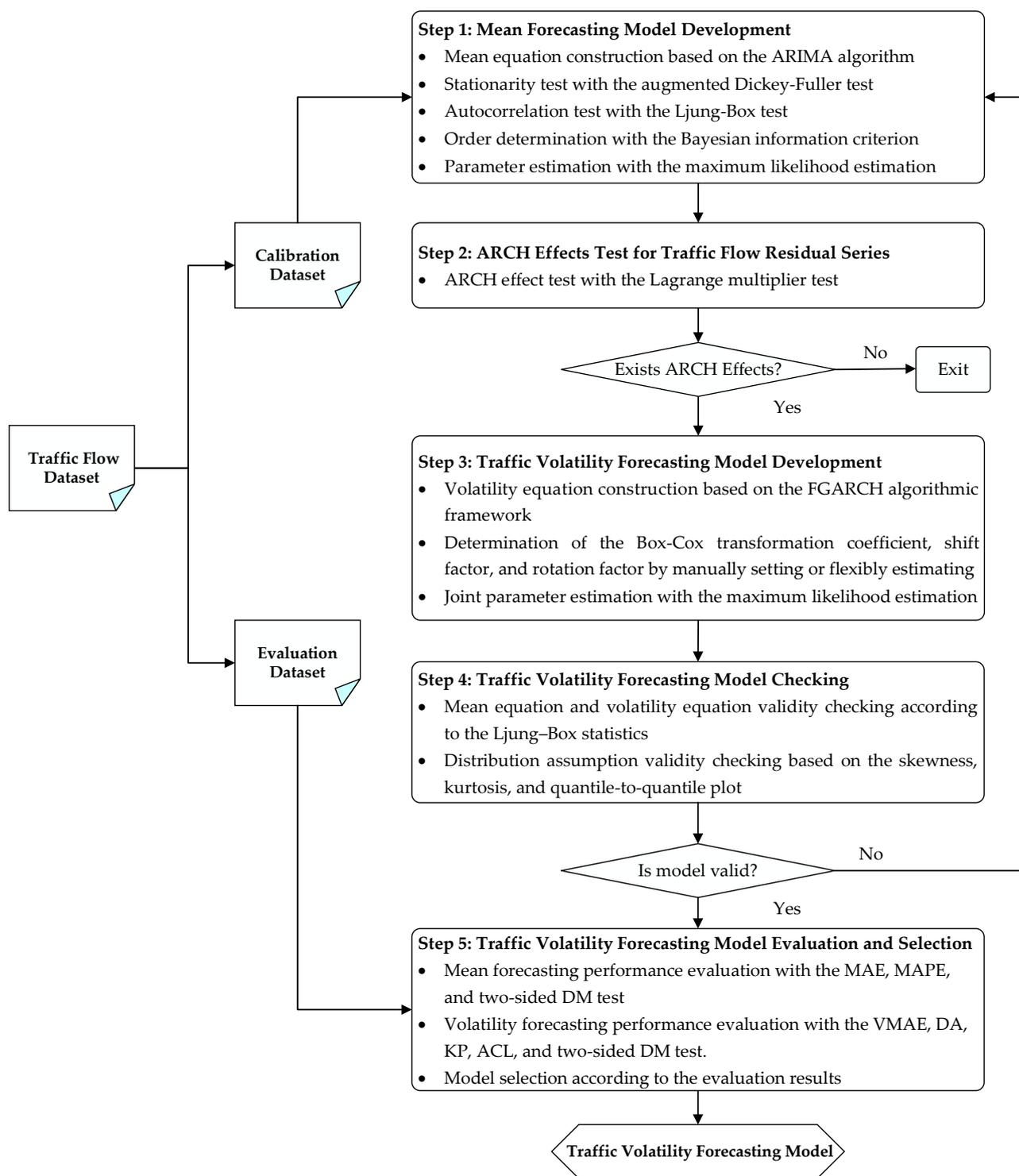


Figure 1. The proposed framework for traffic volatility forecasting.

Generally, the framework can be systematically deconstructed into five steps. First, a mean forecasting model is built based on a calibration dataset and the ARIMA algorithm. To achieve satisfactory mean forecasts, the stationarity test and autocorrelation test are conducted using the augmented Dickey–Fuller (ADF) test [42] and the Ljung–Box test [43], respectively, and the orders of the ARIMA model are optimally identified with a step-wise algorithm [44] based on the Bayesian information criterion (BIC) [45]. Second, conditional heteroscedasticity, also known as the ARCH effect, is checked for the residual series obtained according to the mean equation. To achieve this purpose, the Lagrange

multiplier test [26] is carried out. Third, if the ARCH effects are statistically significant, multiple traffic volatility models based on the FGARCH algorithmic framework are established to capture the symmetric or asymmetric effects in traffic volatility, while a joint estimation of the mean and volatility equations is performed to determine the model parameters using the maximum likelihood estimation (MLE) [46] method. Fourth, the developed models are checked to ensure the validity of the mean equation, volatility equation, as well as the distribution assumption. Finally, the traffic volatility forecasting models are evaluated in terms of performance measures of both mean forecasting and volatility forecasting, and the best model is selected for different application scenarios. In what follows, each step of the proposed framework is described in more details.

### 3.1. Mean Forecasting Model Development

The purpose of developing the mean forecasting model is to remove the sample mean from the traffic flow data and obtain the residual series for further traffic volatility modeling. One of the most widely used mean forecasting models, namely ARIMA [19,20,27,31], can be applied to fit the mean values of the traffic flow time series. Note that if the modeling traffic flow data contains the seasonal components, the seasonal ARIMA (SARIMA) [3,4] model can be constructed to ensure good mean forecasting performance. In this study, we only use one day's data to calibrate the model, where the seasonality is not a typical characteristic of the traffic flow time series. As a result, the ARIMA model is established, which is mathematically expressed as

$$\phi_p(B)(1-B)^d x_t = \theta_q(B)\varepsilon_t, \quad (1)$$

where  $x_t$  is the traffic variable at time  $t$ ,  $B$  is the backshift operator,  $d$  is the differencing order,  $p$  is the autoregressive polynomial order,  $q$  is the moving average polynomial order,  $\phi$  and  $\theta$  are the autoregressive and moving average parameters, and  $\varepsilon_t$  is the error term. After the specification of the mean equation, four further subprocedures need to be carried out to establish the qualified mean forecasting model.

First, the stationarity of the traffic flow time series needs to be checked, which is an important assumption in modeling the ARIMA model. The ADF test [42] is a unit root test, which can be utilized to check the stationarity of a time series. The presence of a unit root implies the analyzed time series is non-stationary and the number of unit roots corresponds to the number of differencing operations required to make the series stationary. For the traffic flow time series  $\{x_t\}$ , to verify the existence of a unit root in an AR ( $p$ ), we can perform the test  $H_0: \beta = 1$  versus  $H_1: \beta < 1$  with an ADF regression

$$x_{\square} = c_t + \beta x_{t-1} + \phi_1 \Delta x_{t-1} + \dots + \phi_{p-1} \Delta x_{t-(p-1)} + \varepsilon_t, \quad (2)$$

where  $c_{\square}$  is a deterministic function of the time index  $t$ , and can be a constant or a linear function of  $t$ ,  $\phi_1, \dots, \phi_{p-1}$  are the autoregressive coefficients,  $\Delta x_{\square} = x_j - x_{j-1}$  is the differenced series of  $x_{\square}$ ,  $\beta$  is the coefficient of  $x_{\square}$ . Equation (2) can also be rewritten as

$$\Delta x_{\square} = c_t + \beta_c x_{t-1} + \phi_1 \Delta x_{t-1} + \dots + \phi_{p-1} \Delta x_{t-(p-1)} + \varepsilon_t, \quad (3)$$

where  $\beta_{\square} = \beta - 1$  and the equivalent hypothesis can be replaced as  $H_0: \beta_c = 0$  versus  $H_1: \beta_c < 0$ . If we accept  $H_0$  means the traffic flow series needs to be differenced to make it stationary, otherwise the stationary condition is satisfied.

Second, the autocorrelation of the traffic flow time series is tested with the Ljung–Box test [43]. As a linear time series model, ARIMA is characterized by its autocorrelation function (ACF), and its modeling process makes use of the sample ACF to specify the forecasting model that can capture the dynamic dependence of the data. In view of this, it is necessary to check whether several autocorrelations of the series  $\{x_t\}$  are 0. The Ljung–Box test is a popular statistical means to achieve the above goal. The test statistic is given as

$$Q(m) = n(n + 2) \sum_{i=1}^m \frac{\hat{\rho}_i^2}{n-i} \tag{4}$$

where  $m$  is the autoregressive lag order of the series  $\{x_t\}$  and can be commonly set as  $\ln(n)$ ,  $n$  is the sample size,  $\hat{\rho}_i$  is the sample ACF of  $\{x_t\}$  and a biased estimate of  $\rho_{\square}$  that is the lag- $i$  autocorrelation of  $\{x_t\}$ . The null hypothesis is  $H_0: \rho_1 = \dots = \rho_m = 0$  and the alternative hypothesis is  $H_1: \rho_i \neq 0$  for  $i \in \{1, \dots, m\}$ . If  $Q(m) > \chi_{\square}^2(\alpha)$ ,  $H_0$  will be rejected, where  $\chi_{\square}^2(\alpha)$  denotes the  $100(1 - \alpha)$ th percentile of a chi-squared distribution with  $m$  degrees of freedom.

Third, a step-wise algorithm proposed by Hyndman and Khandakar [44] is introduced to determine the autoregressive order  $p$  and moving average order  $q$ . One of the advantages of the algorithm is that it is able to automatically identify the appropriate orders of the mean equation and guarantee a valid model. This advantage is significant since the order  $p$  and order  $q$  are often application-dependent, implying the modeling process may be labor-intensive or time-consuming if we determine the orders manually with site-to-site traffic flow data. In this study, the BIC [45] is adopted to measure the quality of each candidate model during the step-wise searching procedure and is defined as

$$BIC(\tau) = \ln(\tilde{\sigma}_{\square}^2) + \frac{\tau \ln(n)}{n}, \tag{5}$$

where  $\tilde{\sigma}_{\square}^2$  is the maximum likelihood estimate of  $\sigma_{\square}^2$  that is the variance of  $\{\varepsilon_{\square}\}$ ,  $\ln(n)$  is the penalty of each parameter. The selection rule is to compute  $BIC(\tau)$  for  $\tau = 0, \dots, \rho$ , where  $\rho$  is a prespecified positive integer, and select the value of  $\tau$  as the order  $p$  or  $q$  that has the minimum BIC value.

Once the orders of the ARIMA are optimally determined, the last subprocedure of building the mean forecasting model is to estimate the parameters  $\phi$  and  $\theta$  with the MLE method. Note that this subprocedure is optional since we can estimate the parameters of the mean equation and volatility equation jointly. In this study, we conduct this subprocedure because we will test whether the joint estimation procedure can indeed improve the point forecasting accuracy. Interested readers can refer to [47] for more details.

### 3.2. ARCH Effects Test for Traffic Flow Residual Series

Once the mean forecasting model is established, the residual series can be computed according to the observations and mean forecasts. The conditional heteroscedasticity of the residual series, also known ARCH effects, needs to be carefully checked to model the traffic volatility. In view of this, the Lagrange multiplier test [26] is introduced to check whether there are ARCH effects in the traffic flow residual series. The Lagrange multiplier test is to conduct testing in the following regression:

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_m \varepsilon_{t-m}^2 + e_t, \quad t = m + 1, \dots, n, \tag{6}$$

where  $\varepsilon_{\square}$  is the residual of the mean equation at time  $t$ ,  $e_t$  denotes the regression error term. More specifically, we define the null hypothesis as  $H_0: \alpha_1 = \dots = \alpha_m = 0$  against the alternative hypothesis as  $H_1: \alpha_i \neq 0$  for some  $i$  between 1 and  $m$ . Denote  $SSR_0 = \sum_{t=m+1}^n (\varepsilon_t^2 - \bar{\omega})^2$ , where  $\bar{\omega} = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2$  is the sample mean of  $\varepsilon_{\square}^2$ , and  $SSR_1 = \sum_{t=m+1}^n \hat{e}_t^2$ , where  $\hat{e}_{\square}$  is the least squares residual of the prior linear regression. Then we have

$$F = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(n - 2m - 1)} \tag{7}$$

which follows an  $F$  distribution with degrees of freedom  $m$  and  $n - 2m - 1$  under  $H_0$ . When  $n$  is sufficiently large,  $mF$  can be used as the test statistic, which is asymptotically a chi-squared distribution with  $m$  degrees of freedom under the null hypothesis. Let  $\chi_m^2(\alpha)$  be the upper  $100(1 - \alpha)$ th percentile of  $\chi_m^2$ . If  $mF > \chi_m^2(\alpha)$ , we reject the null hypothesis and the residual series is heteroscedastic; otherwise, it does not have any ARCH effect. In this study, if the  $P$ -values computed from the  $F$  statistic are less than the

significant level  $\alpha = 0.05$ , a proper volatility model needs to be established to describe the evolution of the traffic flow volatility.

### 3.3. Traffic Volatility Forecasting Model Development

When the significant ARCH effects are checked, the subsequent step is to capture the effects with a proper volatility model. The ARCH model proposed by Engle [26] provides the first systematic framework for volatility modeling. Although owning some advantages for volatility modeling, the ARCH model also has some weaknesses. For example, the ARCH model is rather restrictive and often imposes complicated constraints with the increase of the orders, resulting in the limited applicability of the model. The GARCH model, first introduced by Bollerslev [22], is a useful extension of the ARCH model and can provide a simpler parametric function to describe the volatility evolution. It should be indicated that both models respond equally to the positive and negative shocks [47], and hence cannot be utilized to describe the asymmetric property of the volatility, a common phenomenon in a vast number of application areas.

To deal with the above issue, various GARCH-type models have been presented. Among the GARCH-type models for short-term traffic flow forecasting, the GARCH (1,1) model has been verified as the most adequate and simple representation of traffic volatility [11,18,31]. Therefore, the omnibus FGARCH (1,1) based volatility equation is constructed in this study for short-term traffic volatility forecasting, which is mathematically expressed as

$$\begin{cases} \varepsilon_{\square} = z_t \sigma_t \\ z_t \sim IIN(0,1) \\ \sigma_t^\lambda = \xi + \delta \sigma_{t-1}^\lambda f(\varepsilon_t)^\lambda + \eta \sigma_{t-1}^\lambda \\ f(\varepsilon_t) = |\varepsilon_t - b| - c(\varepsilon_t - b) \end{cases} \quad (8)$$

where  $\varepsilon_{\square}$  is the error term of the mean equation,  $\sigma_{\square}$  is the time-variant conditional standard deviation to reflect the traffic volatility, random variable  $z_{\square}$  is a strong white noise process,  $\xi$ ,  $\delta$ , and  $\eta$  are the parameters that need to be estimated,  $\lambda$  is the Box-Cox transformation coefficient,  $f(\varepsilon_{\square}) = |\varepsilon_t - b| - c(\varepsilon_t - b)$  is the shifted and rotated absolute value function, in which the parameter  $b$  controls the shift effect and the parameter  $c$  governs the rotation effect.

The parameter  $\lambda$  can be obtained from Box-Cox transformation for the conditional standard deviation. Therefore, the relationship between the conditional variances and all shocks in the past may be more flexible than the linear relationship. As indicated by Ding et al. [48], a better estimate of the power transformation coefficient  $\lambda$  can help maintain the long memory property of the volatility. In the standard GARCH model,  $\lambda$  is fixed as 2, which only describes the short memory property of volatility. The shift factor  $b$  and the rotation factor  $c$  are introduced to model the asymmetric impact upon volatility exerted by positive and negative shocks—called the leverage effect in the financial markets. Small shocks are dominated by the shift effect, while the rotation is more important for large shocks [21]. As Helbing et al. [40] demonstrated, there are intrinsic similarities between market players and traffic network users, implying that the shift and rotation effects of the asymmetric property that have been recognized in financial volatility may also exist in traffic volatility.

The above volatility equation has a flexible form for modeling volatility by controlling the parameters  $\lambda$ ,  $b$ , and  $c$ . For instance, we can derive the standard GARCH model by restricting the parameters  $\lambda = 2$  and  $b = c = 0$ . That is, the standard GARCH model imposes restrictive nonnegative constraints on the parameters and also restricts a linear relationship between conditional variances and all shocks in the past. Moreover, the restriction  $b = c = 0$  makes the standard GARCH model a symmetric model, in which both the shift and rotation effects are not allowed to describe the asymmetric property of volatility evolution. Similar restrictions can be seen in other GARCH-type models such as  $\lambda = 0$ ,  $b = 0$  in the EGARCH model, and  $b = 0$  and  $|c| \leq 1$  in the APARCH and TGARCH

models. By manually setting up or freely estimating part or all of the three parameters, we can further obtain multiple models that can capture different traffic flow volatility characteristics. Interested readers can refer to Hentschel [21] for the detailed restrictions of the three parameters in the FGARCH algorithmic framework. Overall, by constructing the FGARCH-based volatility equation, we can implement multiple traffic volatility forecasting models within a unifying framework, and are herein, able to provide a more appropriate and flexible form for explicitly describing the evolution of volatility and to fully investigate the feasibility of the models in different situations. The parameters of the developed traffic volatility forecasting models are finally determined by using the MLE method [46].

### 3.4. Traffic Volatility Forecasting Model Checking

The adequacy of the fitted models needs to be examined carefully before traffic volatility forecasting. More specifically, we need to check the validity of the established mean equation and volatility equation. The validity of the mean equation can be examined by applying the Ljung–Box test to the standardized residual series  $\{\varepsilon_t/\sigma_t\}$  and the validity of the volatility equation can be checked by applying the Ljung–Box test to the squared standardized residual series  $\{(\varepsilon_t/\sigma_t)^2\}$ . When the above  $P$ -values of the Ljung–Box statistic are smaller than  $\chi_m^2(\alpha)$ , the traffic volatility forecasting model can reduce the heteroscedasticity in the residual series and capture the statistical characteristics of traffic volatility well. Moreover, the skewness, kurtosis, and quantile-to-quantile plot of the standardized residual series can be applied to examine the validity of the distribution assumption [47].

### 3.5. Traffic Volatility Forecasting Model Evaluation and Selection

The last step of the framework is to evaluate the performance of the developed models and select the most effective one for further traffic volatility forecasting according to the evaluation results.

For traffic flow mean forecasting performance evaluation, two typical measures, i.e., the mean absolute error (MAE) and the mean absolute percentage error (MAPE), are selected, which are defined as follows.

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |x_t - \hat{x}_t| \quad (9)$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (10)$$

In the above equations,  $x_t$  is the observed value of the traffic variable at time  $t$ ,  $\hat{x}_t$  is the forecast value of the traffic variable at time  $t$ , and  $n$  is the number of the tested data samples. To check the difference of the MAE measures of the two competing models, the two-sided DM test statistic [49] is adopted.

With regard to traffic volatility forecasting performance evaluation, four classical measures, including the kickoff percentage (KP), the average confidence length (ACL), the volatility MAE (VMAE), and the directional accuracy, are employed. For short-term traffic flow forecasting, the KP measure indicates the probability that future observations do not lie in the forecasted confidence intervals when given the evaluation sample size. Theoretically, the better the performance is, the closer the KP is at a given significant level  $\alpha$  ( $\alpha = 0.05$  in this study) [19]. The ACL measure quantifies the average length of the forecasted confidence interval. To reduce forecasting uncertainties, the forecasted confidence interval should be reasonably as narrow as possible. According to Chen et al. [50], VMAE measures the average magnitude of volatility forecasting error, and DA measures the correctness of the turning point forecasts that gives a rough indication of the average direction of the forecast volatility. The four measures are defined as follows.

$$KP = \frac{1}{n} \sum_{t=1}^n K_t \times 100\% \quad (11)$$

$$\begin{aligned} ACL &= \frac{1}{n} \sum_{t=1}^n CL_t \\ &= \frac{1}{n} \sum_{t=1}^n (\hat{U}(x_t) - \hat{L}(x_t)) \end{aligned} \quad (12)$$

$$VMAE = \frac{1}{n} \sum_{t=1}^n |u_t^2 - \hat{u}_t^2| \quad (13)$$

$$DA = \frac{1}{n-1} \sum_{t=2}^n a_t \quad (14)$$

$$a_t = \begin{cases} 1 & (u_t^2 - u_{t-1}^2)(\hat{u}_t^2 - \hat{u}_{t-1}^2) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

In the above equations,  $K_t$  is a two-valued variable,  $K_t = 0$  if  $x_t \in [\hat{L}(x_t), \hat{U}(x_t)]$ ; otherwise,  $K_t = 1$ .  $\hat{L}(x_t)$  and  $\hat{U}(x_t)$  are the lower and upper bounds of the forecasted confidence interval at time  $t$ .  $CL_t$  is the forecasted confidence length at time  $t$ .  $u_t$  and  $\hat{u}_t$  represent the observed volatility and forecasted volatility, respectively. As the observed volatility is unknown in reality,  $u_t^2 = (x_t - \bar{x})^2$  can be used as the surrogate [50], where  $\bar{x}$  is the mean value of the tested data samples. The forecasted volatility is quantified as  $\hat{u}_t^2 = (x_t - \hat{x}_t)^2$ .

## 4. Experimental Analysis

### 4.1. Data Description

The proposed traffic volatility forecasting framework and developed models were tested using real traffic speed data collected from urban roads of Kunshan city, China and freeway segments of the San Diego Region, USA [51]. To investigate the performance of the models under different traffic conditions, two kinds of datasets were constructed. One consists of the datasets collected under normal traffic conditions, while the other includes the datasets collected under incident conditions. The former datasets were collected at 35 stations on 5 types of roads, namely major arterials, minor arterials, branches, expressway, and freeway. For each station, we collected data in four days. Two of them are the weekdays, and the other two are weekends. For the latter, the traffic flow data was acquired from the freeway monitoring stations associated with 12 incidents occurred in different days. The traffic speed data of each station was aggregated at 5-min intervals and divided into a calibration dataset and an evaluation dataset. The calibration dataset was used to build the traffic volatility model, while the evaluation dataset was utilized to test the performance of the calibrated model. More detailed information on the used data is described in Appendix A Tables A1 and A2. It is noted that the missing, erroneous, and suspicious data is less than 3% for each monitoring station during the data collection periods, ensuring that the selected traffic datasets are suitable for model calibration and evaluation.

### 4.2. Performance Evaluation of Mean Forecasting Models

To evaluate the mean forecasting performance of the developed models, the MAE and its two-sided DM test statistics were computed and compared. In this study, the parameters of the mean forecasting model can be estimated in two ways—namely separate estimation and joint estimation. The former is implemented based on a separate ARIMA modeling process, while the latter is achieved by estimating the parameters of the mean equation and volatility equation jointly. The two ways may lead to different mean forecasting performance. Therefore, it is necessary to investigate the performance difference between the two ways.

Table 1 shows the performance comparisons between the separate estimation and joint estimation on the evaluation datasets. The first column of the table indicates the traffic flow patterns considered in this research, including normal traffic patterns on weekdays, normal traffic patterns on weekends, and incident traffic patterns. The second column depicts the comparison results. The label “Win” means the joint estimation way outperforms the separate estimation way. The label “Lose” means the joint estimation way is worse than the separate estimation way. The label “Tie” means the two ways have identical or similar performance. The digital number in the table cell represents the number of stations associated with the label in the corresponding row. The percentage in the parentheses indicates the proportion of the associated label.

As can be seen from the table, under normal traffic conditions, the two estimation ways show very similar performance, especially on weekends. However, under incident traffic conditions, there were significant differences between the performance of the two estimation ways on nearly half of the evaluation stations. More specifically, the separate estimation way achieved better performance than the joint estimation way. This implies that it is better to first estimate the parameters of the mean equation and further use the residual time series to estimate the parameters of the volatility equation when modeling traffic volatility under incident traffic conditions.

**Table 1.** Mean forecasting performance comparisons of separate estimation and joint estimation.

Pattern	Result	Joint Estimation vs. Separate Estimation					
		S	T	N	NA	GJR	F
Normal (Weekday)	Win	3 (8.57%)	4 (11.43%)	4 (11.43%)	3 (8.57%)	3 (8.57%)	5 (14.29%)
		3 (8.57%)	7 (20.00%)	5 (14.29%)	3 (8.57%)	7 (20.00%)	4 (11.43%)
	Lose	29 (82.86%)	24 (68.57%)	26 (74.29%)	29 (82.86%)	25 (71.43%)	26 (74.29%)
		Total	35	35	35	35	35
Normal (Weekend)	Win	3 (8.57%)	3 (8.57%)	2 (5.71%)	2 (5.71%)	2 (5.71%)	4 (11.43%)
		3 (8.57%)	1 (2.86%)	1 (2.86%)	3 (8.57%)	1 (2.86%)	1 (2.86%)
	Lose	29 (82.86%)	31 (88.57%)	32 (91.43%)	30 (85.71%)	32 (91.43%)	30 (85.71%)
		Total	35	35	35	35	35
Incident	Win	2 (16.67%)	3 (25.00%)	3 (25.00%)	2 (16.67%)	2 (16.67%)	2 (16.67%)
		5 (41.67%)	3 (25.00%)	3 (25.00%)	3 (25.00%)	5 (41.67%)	5 (41.67%)
	Lose	5 (41.67%)	6 (50.00%)	6 (50.00%)	7 (58.33%)	5 (41.67%)	5 (41.67%)
		Total	12	12	12	12	12

S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

Traffic on different types of roads exhibits distinct fluctuation patterns. To properly evaluate the performance of the established models under different volatility levels, we grouped the stations with different road types according to the calculated VMAE measure, as illustrated in Table 2. As seen, for normal conditions on weekdays, the traffic on minor arterial and branch show significant fluctuations and had a high volatility level. The major arterial and freeway possess medium volatility level, while the expressway has the lowest volatility. Different from the normal conditions on weekdays, the daily traffic on weekends passed through the freeway appears to have very low fluctuations and was

hence put into the low volatility level group. The traffic flow on freeway segments shows medium-level volatility.

**Table 2.** Volatility level of the traffic speeds collected from different types of roads.

Road Type	Volatility under Normal Conditions (Weekday)	Volatility under Normal Conditions (Weekend)	Volatility under Incident Conditions
Major Arterial	126.529 (Medium)	96.567 (Medium)	
Minor Arterial	158.048 (High)	164.377 (High)	
Branch	135.374 (High)	129.801 (High)	
Expressway	6.014 (Low)	5.232 (Low)	
Freeway	129.723 (Medium)	5.691 (Low)	54.772 (Medium)

Table 3 provides the average MAE value and its DM test rank of the developed models on the evaluation datasets. As mentioned above, the stations were categorized into three groups (H-high volatility level; M-medium volatility level; L-low volatility level) according to the volatility level. The average MAE value was calculated as the mean value of all stations in the same group. The DM test rank was computed based on the following procedure: First, for each station in the same volatility level group, we compare the performance of two models according to the DM statistic and *p* value. If the *p* value is less than the significant level (0.05 in this study), it means one of the two comparative models is significantly better or worse than the other model. By repeating the above comparison process, we can obtain the rank of each model. Next, the ranks of each model are averaged for all stations in the same volatility level group. The smaller the rank is, the better performance the model possesses.

The best rank is labeled with the red color, while the worst rank is labeled with the blue color. Note that when the DM test ranks of two models are same, the best model is further determined by the average MAE values. As indicated by the average MAE values, the developed six models show very similar mean forecasting performance. By further going through the DM test ranks, we find that the NGARCH has better performance when the volatility levels are high and medium. Comparatively, the TGARCH shows worse performance in the same situations. It is a bit surprising that the TGARCH model shows advantages over other models under incident conditions. We can also see that the freely estimated FGARCH model takes the most advantage at the high volatility level on weekdays.

**Table 3.** Average MAE value and its DM test rank of the developed models on evaluation datasets.

MAE	GARCH Type	Normal (Weekday)			Normal (Weekend)			Incident
		H	M	L	H	M	L	
Value	S	4.241	2.867	0.980	4.249	3.422	0.757	1.222
	T	4.251	3.461	0.998	4.244	3.426	0.771	1.149
	N	4.246	2.705	0.982	4.244	3.422	0.760	1.165
	NA	4.241	2.868	1.010	4.250	3.424	0.760	1.208
	GJR	4.238	2.877	1.003	4.247	3.426	0.856	1.222
	F	4.322	2.728	0.987	4.257	3.427	0.856	1.169
Rank	S	1.643	1.400	1.000	1.714	1.000	2.417	2.917
	T	2.143	1.733	1.500	1.786	1.556	2.417	2.417
	N	1.643	1.067	1.833	1.286	1.000	2.250	2.833
	NA	1.786	1.467	2.500	1.857	1.000	1.750	2.667
	GJR	1.714	1.733	1.667	1.286	1.000	2.500	3.250
	F	1.429	1.267	2.167	1.714	1.000	1.833	3.083

S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

Table 4 gives the average MAPE value and its rank for each developed model. The computing procedure is similar with that of the MAE measure. From the evaluation results, we can obtain a similar observation, further confirming the conclusions achieved above.

**Table 4.** Average MAPE value and its rank of the developed models on evaluation datasets.

MAPE	GARCH Type	Normal (Weekday)			Normal (Weekend)			Incident
		H	M	L	H	M	L	
Value	S	14.865	9.231	1.392	14.456	8.316	1.095	2.464
	T	14.924	10.133	1.425	14.478	8.324	1.114	2.383
	N	14.881	8.988	1.397	14.429	8.313	1.099	2.383
	NA	14.884	9.220	1.440	14.504	8.316	1.097	2.431
	GJR	14.819	9.271	1.430	14.441	8.318	1.234	2.458
	F	15.121	9.039	1.405	14.528	8.314	1.235	2.384
Rank	S	2.929	2.667	1.333	3.214	2.111	2.750	2.500
	T	3.143	3.267	1.667	2.786	2.889	2.167	2.417
	N	2.714	2.933	2.000	2.500	1.444	2.500	2.333
	NA	2.429	1.933	3.000	3.357	1.778	1.500	2.417
	GJR	2.357	2.600	2.000	2.714	2.333	2.167	3.083
	F	2.286	3.067	2.667	3.000	2.222	2.083	3.000

S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

### 4.3. Performance Evaluation of Volatility Forecasting Models

To effectively evaluate the volatility forecasting performance of the developed models, the VMAE, DA, KP, and ACL measures were calculated. The four measures quantified the model performance from different perspectives. The VMAE measures the accuracy of traffic volatility forecasting. The DA gives the volatility direction accuracy of the forecasting models. KP and ACL ensure the forecasts fall into the predicted confidence interval as much as possible when the interval length is short enough. Considering the inconsistency of the measures in some cases, we also computed the ranks of the models for each measure to make a fair comparison.

Table 5 lists the average VMAE value and its corresponding DM test rank of the developed models. As can be seen, the volatility forecasting performance is quite similar for the six models. However, their ranks indicate that different models usually show distinct advantages compared to different traffic speed datasets. As indicated by Hentschel [21], the GARCH-type models developed in this study show similar forecasting results on the Daily U.S. stock dataset due to the fact that they are derived from the same modeling framework. Our experimental results also point to a similar conclusion, although the models indeed show distinct forecasting behaviors on several datasets. For instance, Table 6 and Figure 2 depict the distinct performance of the models on the incident dataset I818\_1118078.

As seen in Table 3, the estimated parameters of the mean equation of the models appear to be very similar because the orders of the equation are the same. For the volatility equation, it is obvious that the estimated values of the parameter  $\lambda$  exhibit distinct diversity among different models. The different values of the parameter  $\lambda$  implies that the developed models may capture the conditional variance and respond differently to the past shocks when modeling the same traffic flow series. In addition, both of the freely estimated parameters  $b$  and  $c$  are different from 0, indicating the existence of asymmetric property of traffic speed volatility controlled by the shift and rotation effects at the same time. Moreover, for the FGARCH model, the estimated parameter  $b$  is positive and parameter  $c$  is negative, implying that the asymmetric property of traffic speed volatility at this station may be dominated by the combination of rightward shift effect and

anticlockwise rotation effect, which is consistent with the “news impact curve” introduced by Pagan and Schwert [52].

**Table 5.** Average VMAE value and its DM test rank of the developed models on evaluation datasets.

VMAE	GARCH Type	Normal (Weekday)			Normal (Weekend)			Incident
		H	M	L	H	M	L	
Value	S	152.915	127.988	6.061	154.412	96.573	5.462	58.940
	T	152.703	264.279	6.006	154.398	96.561	5.459	56.690
	N	152.912	111.636	6.061	154.493	96.574	5.457	56.560
	NA	152.909	128.103	6.013	154.168	96.551	5.450	59.080
	GJR	153.084	127.614	6.015	154.407	96.512	5.652	58.970
	F	153.489	111.735	6.051	154.226	96.623	5.754	56.560
Rank	S	2.071	1.667	1.000	1.786	1.556	1.667	1.250
	T	1.500	2.067	1.000	1.929	1.333	1.667	2.170
	N	2.214	1.867	1.000	2.000	1.556	1.583	1.330
	NA	2.071	2.200	1.000	1.429	1.444	1.917	1.670
	GJR	2.214	1.867	1.000	1.929	1.222	1.917	1.670
	F	2.286	2.133	1.000	1.357	1.556	2.750	1.830

S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

**Table 6.** Joint parameter estimation of the mean equation and volatility equation for six developed models with the traffic speed dataset collected at station I818\_1118078 on 10 August 2010.

Mean Equation		Volatility Equation		
ARIMA (2,1,1)	$\phi_1 = 0.454$	GARCH (1,1)	$\xi = 0.024$	$\lambda = 2.00$
	$\phi_2 = 0.001$		$\delta = 0.057$	$b = 0.00$
	$\theta_1 = -0.903$		$\eta = 0.904$	$c = 0.00$
	$\phi_1 = 0.458$	TGARCH (1,1)	$\xi = 0.022$	$\lambda = 1.00$
	$\phi_2 = -0.006$		$\delta = 0.073$	$b = 0.00$
	$\theta_1 = -0.890$		$\eta = 0.916$	$c = 0.32$
	$\phi_1 = 0.457$	NGARCH (1,1)	$\xi = 0.003$	$\lambda = 4.00$
	$\phi_2 = -0.008$		$\delta = 0.000$	$b = 0.00$
	$\theta_1 = -0.905$		$\eta = 0.983$	$c = 0.00$
	$\phi_1 = 0.451$	NAGARCH (1,1)	$\xi = 0.016$	$\lambda = 2.00$
	$\phi_2 = -0.008$		$\delta = 0.057$	$b = 0.49$
	$\theta_1 = -0.896$		$\eta = 0.905$	$c = 0.00$
	$\phi_1 = 0.452$	GJR-GARCH (1,1)	$\xi = 0.018$	$\lambda = 2.00$
	$\phi_2 = 0.001$		$\delta = 0.055$	$b = 0.00$
	$\theta_1 = -0.902$		$\eta = 0.914$	$c = 0.19$
	$\phi_1 = 0.435$	FGARCH (1,1)	$\xi = 0.104$	$\lambda = 0.04$
	$\phi_2 = 0.006$		$\delta = 0.072$	$b = 0.87$
	$\theta_1 = -0.891$		$\eta = 0.825$	$c = -0.58$

Figure 2 depicts the forecasted speed standard deviation curves of the developed models. As seen from the figure, these models respond differently to the fluctuations induced by the incident. The NGARCH model decreases the standard deviation forecasts gradually with the time evolution, while the other five models are all able to capture the incident volatility but respond with different magnitudes. To further check the effectiveness of the models, the observation, mean forecasts, as well as the forecast confidence intervals are given in Appendix A Figure A1. It can easily be seen that the FGARCH exhibits the best volatility forecasting performance among the six models because it can not only correctly capture the volatility characteristics caused by the incident but also provide more reliable and narrower confidence interval than the others.

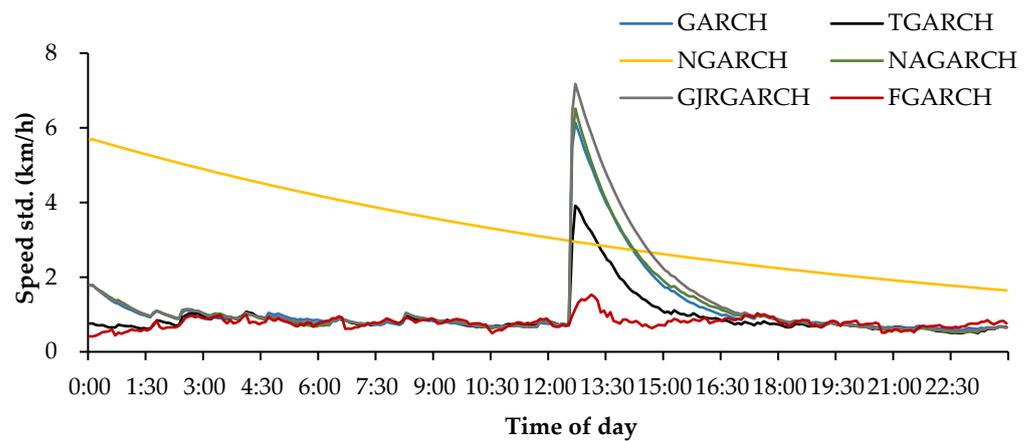


Figure 2. Comparison of forecasted traffic speed standard deviations of six developed models.

The DA measure describes the forecasting directional accuracy of the given model. In view of this, we calculated the average DA value and its rank of the six models, as shown in Table 7. From the table, it is obvious that under normal traffic conditions, the standard GARCH model has the best performance on weekdays and the NAGARCH model shows more advantages over other models. Interestingly, the FGARCH has the best performance under incident conditions while not performing well under normal traffic conditions.

Table 7. Average DA value and its rank of the developed models on evaluation datasets.

DA	GARCH Type	Normal (Weekday)			Normal (Weekend)			Incident
		H	M	L	H	M	L	
Value	S	0.681	0.632	0.853	0.687	0.685	0.711	0.600
	T	0.682	0.625	0.845	0.691	0.687	0.709	0.604
	N	0.679	0.634	0.853	0.685	0.689	0.712	0.602
	NA	0.680	0.634	0.839	0.694	0.688	0.718	0.602
	GJR	0.679	0.636	0.845	0.684	0.688	0.716	0.599
	F	0.669	0.638	0.844	0.688	0.686	0.710	0.604
Rank	S	2.000	2.133	1.333	2.786	2.778	3.167	2.667
	T	2.071	3.333	2.667	2.714	2.556	3.333	2.750
	N	3.500	2.533	1.333	3.000	1.889	3.583	2.500
	NA	2.643	2.867	4.833	2.000	1.778	2.333	2.333
	GJR	2.143	2.667	2.667	3.643	1.889	2.250	2.583
	F	3.643	3.067	2.833	2.714	3.667	3.250	2.333

S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

The absolute deviation of KP from the significance level of 0.05 was calculated to check the validity of the models at each traffic station. The calculation results include the average value and its rank on all stations of the same volatility level group, which are illustrated in Table 8. From the table, we can see that the standard GARCH model and the GJR-GARCH model show better performance than the other four models in terms of KPD, which means the two models exhibit the smallest absolute deviation of KP from the given significant level ( $\alpha = 0.05$ ) and tried to make the confidence interval contain forecasted data samples as much as possible. In contrast, the FGARCH exhibits unsatisfactory performance in most cases.

The average ACL value and the rank of each developed model are presented in Table 9. As can be seen from the table, the overall ACL values of the models show similarities in different situations. Relatively, the standard GARCH model, GJR-GARCH model, and

FGARCH model tend to enlarge the forecasted confidence interval. In contrast, the TGARCH model attempts to give a narrower confidence interval in most cases. It needs to be indicated that it is commonly difficult for a volatility model to achieve the lowest KP and ACL at the same time because the two measures are counterbalanced in describing volatility characteristics.

**Table 8.** Average KPD value and its rank of six developed models on evaluation datasets.

KPD	GARCH Type	Normal (Weekday)			Normal (Weekend)			Incident
		H	M	L	H	M	L	
Value	S	0.010	0.007	0.012	0.011	0.011	0.009	0.008
	T	0.013	0.008	0.018	0.012	0.013	0.011	0.034
	N	0.012	0.007	0.168	0.011	0.014	0.011	0.110
	NA	0.014	0.009	0.016	0.015	0.011	0.011	0.028
	GJR	0.010	0.006	0.014	0.012	0.009	0.012	0.010
	F	0.089	0.077	0.014	0.022	0.019	0.017	0.031
Rank	S	2.286	2.600	2.500	2.429	2.333	2.583	1.583
	T	3.286	3.867	4.000	2.571	3.000	2.750	3.417
	N	2.500	2.800	3.000	2.429	3.000	3.250	4.750
	NA	3.857	2.600	3.333	3.214	2.778	3.417	2.250
	GJR	2.500	2.200	2.833	2.286	1.889	2.917	2.083
	F	3.500	4.200	3.500	4.929	5.000	3.917	3.583

S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

**Table 9.** Average ACL value and its rank of six developed models on evaluation datasets.

ACL	GARCH Type	Normal (Weekday)			Normal (Weekend)			Incident
		H	M	L	H	M	L	
Value	S	21.789	13.995	5.193	21.354	16.974	3.897	6.833
	T	22.423	15.343	5.113	21.137	16.864	3.952	4.867
	N	21.992	13.447	4.245	21.303	17.044	3.846	5.772
	NA	21.749	13.954	5.122	21.576	16.949	3.870	6.324
	GJR	22.024	14.017	5.072	21.388	16.931	4.163	6.783
	F	21.554	13.389	5.313	22.021	16.794	4.113	7.856
Rank	S	3.429	4.067	3.833	3.214	3.222	3.833	3.833
	T	3.571	4.333	2.667	2.500	3.222	2.917	2.250
	N	3.214	2.600	3.500	3.714	4.000	3.250	3.250
	NA	2.786	3.333	3.500	3.500	3.333	2.833	3.583
	GJR	3.857	3.533	2.833	3.786	3.889	3.500	3.750
	F	4.071	2.733	4.167	4.143	2.444	3.667	3.417

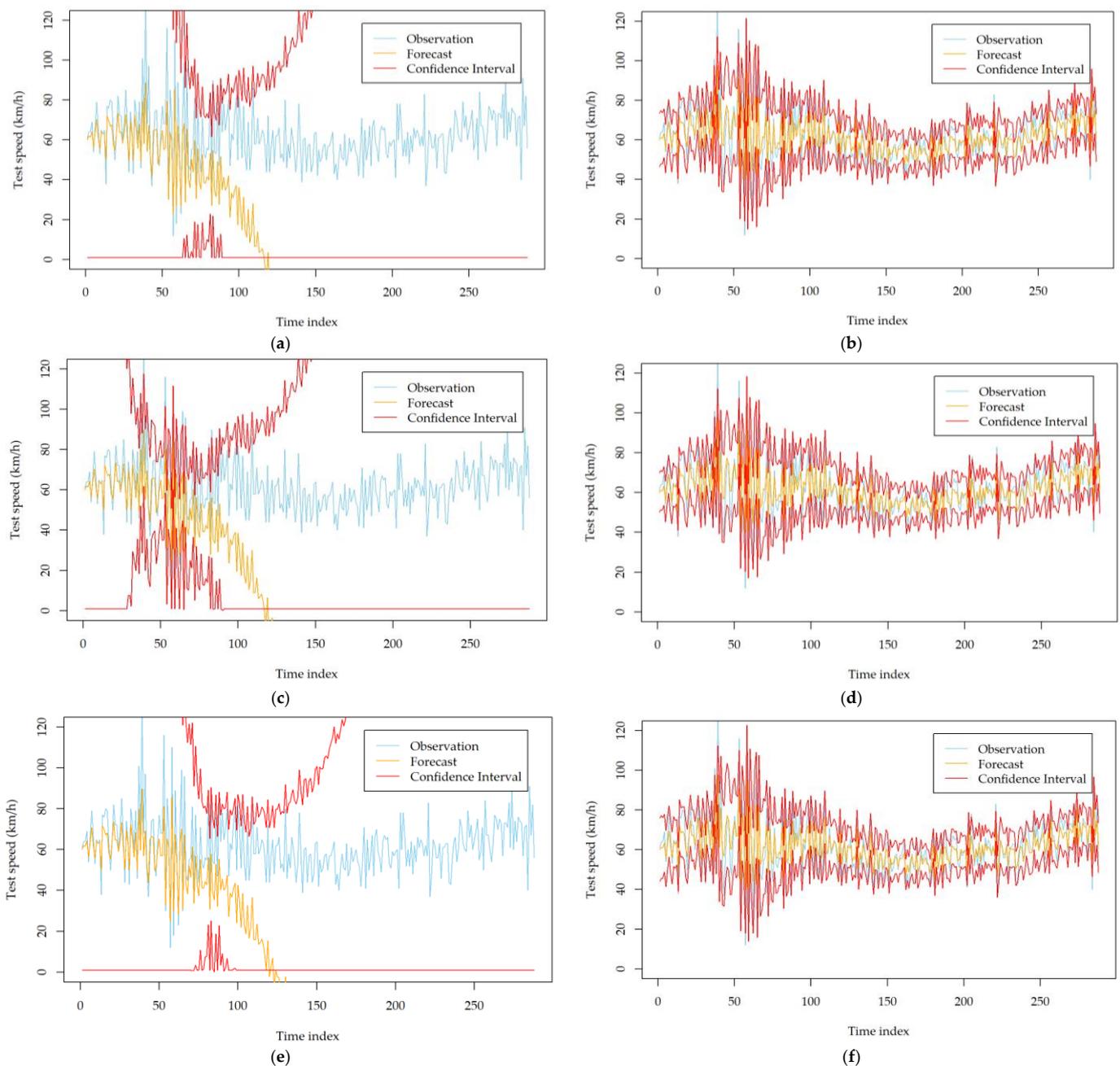
S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

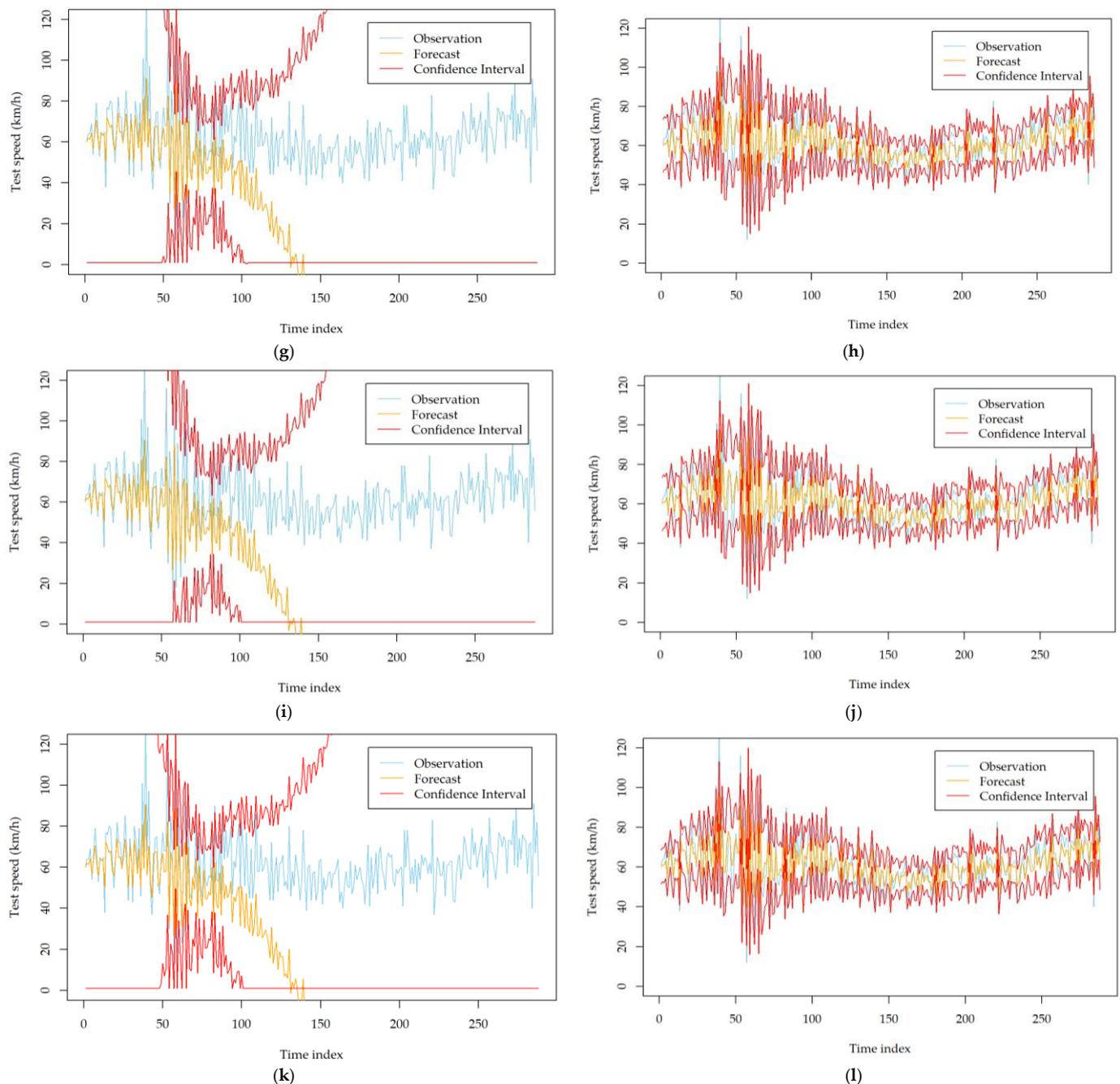
#### 4.4. Impact of Identifying Orders of ARIMA on Forecasting Performance

There are commonly two strategies for identifying the orders of the mean equation of the proposed framework. The first is an automatic identification strategy, which uses the step-wise searching algorithm [44] with an information criteria (e.g., AIC or BIC). The other is a manual identification strategy that finds the optimal orders by manually inspecting the ACF and PACF plots. Each strategy has its own advantages. For example, the automatic identification strategy is more computationally efficient and can be utilized to handle extensive datasets in an automated manner and hence realize significant cost savings, while the manual identification strategy is capable of providing more reliable forecasts in some cases.

To investigate the impact of the two order identification strategies on the forecasting performance of the developed models, we recorded the forecasting results using the two

strategies. After careful inspection and calibration, we found that the two strategies lead to similar forecasting performance in most stations. However, for several specific monitoring stations, they show significant differences. Appendix A Tables A3 and A4 show the comparison results on such kinds of stations. From the tables, we can see that the models established based on the automatic strategy can result in a significantly poor mean forecasting performance and volatility forecasting performance. The MAE and DA deteriorate dramatically for all of the developed models. This observation can be further obtained with Figure 3. The mean forecasts and volatility forecasts bias the true observations significantly, while they are improved evidently after using the manual order identification strategy.





**Figure 3.** Comparison of the volatility forecasting performance of six developed models with different order identification strategies; (a) GARCH (automatic order identification); (b) GARCH (manual order identification); (c) TGARCH (automatic order identification); (d) TGARCH (manual order identification); (e) NGARCH (automatic order identification); (f) NGARCH (manual order identification); (g) NAGARCH (automatic order identification); (h) NAGARCH (manual order identification); (i) GJR-GARCH (automatic order identification); (j) GJR-GARCH (manual order identification); (k) FGARCH (automatic order identification); (l) FGARCH (manual order identification).

## 5. Conclusions

Traffic volatility forecasting plays a key role in traffic reliability or uncertainty analysis. Many GARCH-type models have been developed to capture volatility in the fields of finance and economics. However, the ability of these models to capture traffic volatility has not been fully investigated. In this study, a systematic traffic volatility forecasting framework based on an omnibus family GARCH model is proposed. Under this framework, six volatility models were developed by manually setting up or freely estimating

three key parameters. The performance of the models was comprehensively evaluated and compared by using 47 traffic speed data collected from five kinds of roads under normal traffic conditions and incident conditions. A number of key findings were achieved based on the experimental results, which can provide insights into traffic volatility modeling for researchers and practitioners.

The experimental results in this research reveal the potential of explicitly modeling volatility evolution properties in improving traffic volatility forecasting performance. In the future, more traffic flow datasets could be used to evaluate the effectiveness of the proposed framework. The physical explanations about how the asymmetric property of traffic volatility is associated with various traffic phenomena are also worth exploring.

**Author Contributions:** Conceptualization, J.O., Q.N. and Y.Z.; methodology, J.O., Y.Z. and Q.N.; software, J.O., X.H. and Z.Z.; validation, J.O., X.H., Y.Z., Z.Z. and Q.N.; writing—original draft preparation, J.O., X.H. and Q.N.; funding acquisition, J.O. and Q.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the Humanity and Social Science Youth Foundation of Ministry of Education of China (No. 22YJC630109), the Yangzhou Natural Science Foundation (No. YZ2019079), and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 20KJB580012).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Traffic flow datasets collected under normal traffic conditions.

Station ID	Road Type	Calibration Dataset		Evaluation Dataset	
		Period	Sample Size	Period	Sample Size
MA_1004015	Major Arterial	4 September 2018	288	5 September 2018	288
MA_1004027					
MA_1004028					
MA_1012021					
MA_1014005		8 September 2018	288	9 September 2018	288
MA_1014009					
MA_1014014					
MA_1014018					
MA_1016012					
MI_1003001	Minor Arterial	4 September 2018	288	5 September 2018	288
MI_1004019					
MI_1005012					
MI_1005013					
MI_1005015		8 September 2018	288	9 September 2018	288
MI_1005016					
MI_1005018					
MI_1006019					
MI_1006020					
B_1004017	Branch	4 September 2018	288	5 September 2018	288
B_1013001		8 September 2018	288	9 September 2018	288
B_1013003					

B_1015003					
B_1015005					
E_10131301					
E_10131501					
E_10132601	Expressway	4 September 2018	288	5 September 2018	288
E_10141101		8 September 2018	288	9 September 2018	288
E_10141201					
E_10141401					
F_1108452					
F_1108498					
F_1114254	Freeway	2 March 2010	288	3 March 2010	288
F_1114735		6 March 2010	288	7 March 2010	288
F_1114806					
F_1117755					

Table A2. Traffic flow datasets collected under incident conditions.

Station ID	Road Type	Calibration Dataset		Evaluation Dataset	
		Period	Sample Size	Period	Sample Size
I190_1117774	Freeway	25 July 2010	288	26 July 2010	288
I224_1117755	Freeway	25 July 2010	288	26 July 2010	288
I275_1117835	Freeway	8 April 2010	288	9 April 2010	288
I371_1114806	Freeway	24 September 2010	288	25 September 2010	288
I373_1108711	Freeway	8 October 2010	288	9 October 2010	288
I383_1115355	Freeway	4 April 2010	288	5 April 2010	288
I393_1108615	Freeway	19 June 2010	288	20 June 2010	288
I689_1118537	Freeway	15 July 2010	288	16 July 2010	288
I772_1117835	Freeway	31 July 2010	288	1 August 2010	288
I818_1118078	Freeway	10 August 2010	288	11 August 2010	288
I914_1117809	Freeway	2 September 2010	288	3 September 2010	288
I1039_1118078	Freeway	6 October 2010	288	7 October 2010	288

Table A3. Comparison of the MAE measure of the developed models with different order identification strategies.

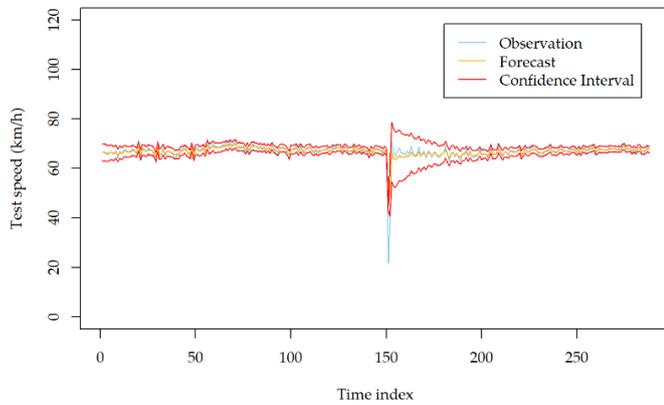
Station	Test Period	Strategy	$(p, d, q)$	MAE (veh/5 min)					
				S	T	N	NA	GJR	F
MA_1014014	9 September 2018	Automatic	(2,1,3)	1297.99	1298.40	962.04	554.66	582.07	522.14
		Manual	(0,1,2)	5.02	5.03	5.02	5.02	5.02	5.04
MI_1005012	9 September 2018	Automatic	(3,1,3)	173.48	3.20	13.22	386.48	9.60	33.73
		Manual	(0,1,2)	3.19	3.20	3.18	3.18	3.19	3.18
E_10141201	5 September 2018	Automatic	(2,0,2)	4.62	8.16	1.91	4.31	4.25	1.96
		Manual	(0,1,1)	1.05	1.18	1.05	1.21	1.20	1.06
I818_1118078	11 August 2010	Automatic	(2,0,1)	2.23	6.25	1.45	2.24	2.25	125.03
		Manual	(0,1,1)	0.76	0.76	0.75	0.76	0.76	0.76

S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).

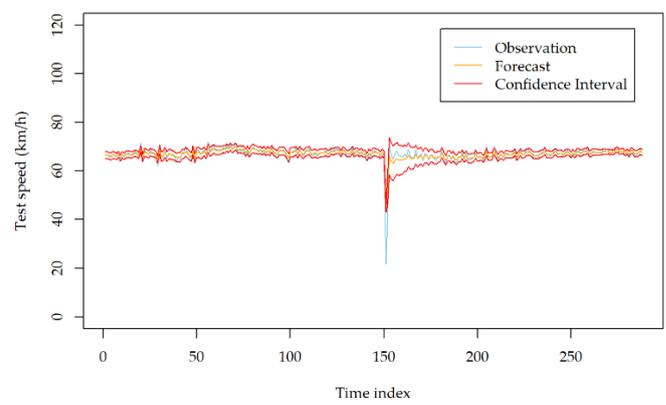
**Table A4.** Comparison of the DA measure of the developed models with different order identification strategies.

Station	Test Period	Strategy	$(p, d, q)$	DA					
				S	T	N	NA	GJR	F
MA_1014014	9 September 2018	Automatic	(2,1,3)	0.5505	0.5679	0.5784	0.5505	0.5540	0.5505
		Manual	(0,1,2)	0.7666	0.7561	0.7700	0.7666	0.7666	0.7491
MI_1005012	9 September 2018	Automatic	(3,1,3)	0.4913	0.5923	0.5366	0.4948	0.5575	0.5331
		Manual	(0,1,2)	0.5854	0.5854	0.5958	0.5854	0.5889	0.5854
E_10141201	5 September 2018	Automatic	(2,0,2)	0.7944	0.7561	0.8118	0.8014	0.7944	0.7979
		Manual	(0,1,1)	0.8328	0.8014	0.8328	0.7979	0.8014	0.8328
I818_1118078	11 August 2010	Automatic	(2,0,1)	0.6028	0.6411	0.5923	0.6063	0.6028	0.5819
		Manual	(0,1,1)	0.7387	0.7387	0.7422	0.7352	0.7387	0.7282

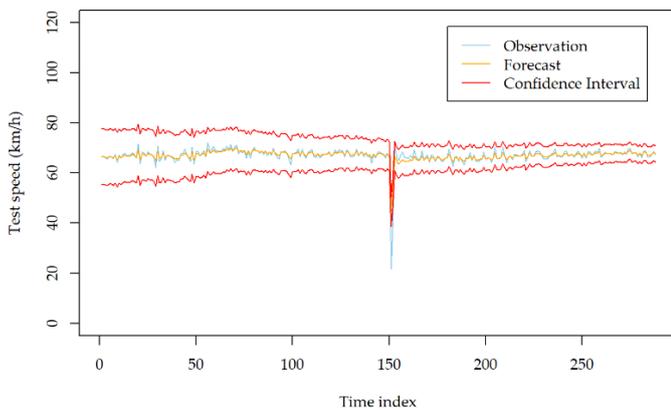
S (Standard GARCH), T (TGARCH), N (NGARCH), NA (NAGARCH), GJR (GJR-GARCH), F (FGARCH).



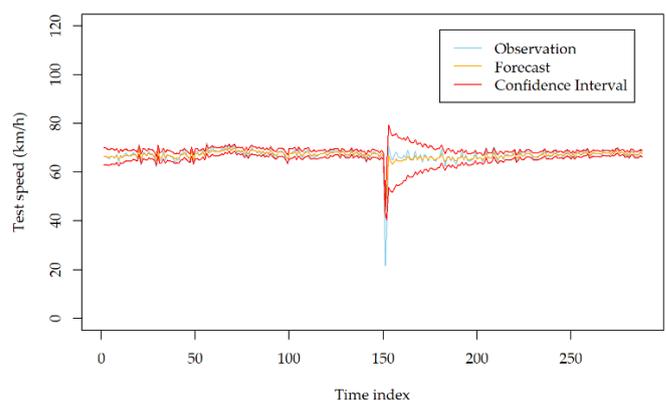
(a)



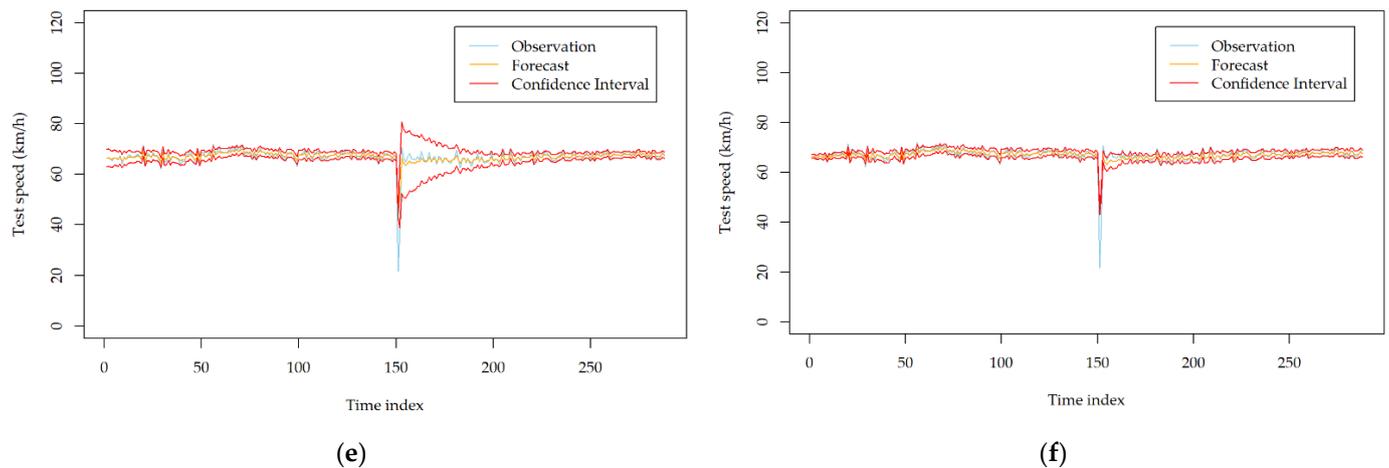
(b)



(c)



(d)



**Figure A1.** Volatility forecasting results of the developed models with the traffic speed dataset collected at station I818\_1118078 on 10 August 2010; (a) GARCH; (b) TGARCH; (c) NGARCH; (d) NA-GARCH; (e) GJR-GARCH; (f) FGARCH.

## References

- Ishak, S.; Al-Deek, H. Performance Evaluation of Short-Term Time-Series Traffic Prediction Model. *J. Transp. Eng.* **2002**, *128*, 490–498.
- Ishak, S.; Al-Deek, H. Statistical Evaluation of Interstate 4 Traffic Prediction System. *Transp. Res. Rec. J. Transp. Res. Board* **2003**, *1856*, 16–24.
- Williams, B.M.; Hoel, L.A. Modeling and Forecasting Vehicular Traffic Flow as A Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *J. Transp. Eng.* **2003**, *129*, 664–672.
- Shekhar, S.; Williams, B.M. Adaptive Seasonal Time Series Models for Forecasting Short-Term Traffic Flow. *Transp. Res. Rec. J. Transp. Res. Board* **2008**, *2024*, 116–125.
- Guo, J.; Xia, J.; Smith, B.L. Kalman Filter Approach to Speed Estimation Using Single Loop Detector Measurements under Congested Conditions. *J. Transp. Eng.* **2009**, *135*, 927–934.
- Gazis, D.; Liu, C. Kalman Filtering Estimation of Traffic Counts for Two Network Links in Tandem. *Transp. Res. Part B Methodol.* **2003**, *37*, 737–745.
- Davis, G.A.; Nihan, N.L. Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *J. Transp. Eng.* **1991**, *117*, 178–188.
- Xia, J.; Chen, M.; Huang, W. A Multistep Corridor Travel-Time Prediction Method Using Presence-Type Vehicle Detector Data. *J. Intell. Transp. Syst.* **2011**, *15*, 104–113.
- Park, B.; Messer, C.J.; Urbanik, T. Short-term Freeway Traffic Volume Forecasting Using Radial Basis Function Neural Network. *Transp. Res. Rec. J. Transp. Res. Board* **1998**, *1651*, 39–47.
- Rilett, L.R.; Park, D. Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks. *Transp. Res. Rec. J. Transp. Res. Board* **2001**, *1752*, 140–147.
- Fei, X.; Zhang, Y.; Liu, K.; Guo, M. Bayesian Dynamic Linear Model with Switching for Real-Time Short-Term Freeway Travel Time Prediction with License Plate Recognition Data. *J. Transp. Eng.* **2013**, *139*, 1058–1067.
- Gu, Y.; Lu, W.; Xu, X.; Qin, L.; Shao, Z.; Zhang, H. An Improved Bayesian Combination Model for Short-Term Traffic Prediction with Deep Learning. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1332–1342.
- Lu, Z.; Xia, J.; Wang, M.; Nie, Q.; Ou, J. Short-Term Traffic Flow Forecasting via Multi-Regime Modeling and Ensemble Learning. *Appl. Sci.* **2020**, *10*, 356.
- Ou, J.; Xia, J.; Wu, Y.; Rao, W. Short-Term Traffic Flow Forecasting for Urban Roads Using Data-Driven Feature Selection Strategy and Bias-Corrected Random Forests. *Transp. Res. Rec. J. Transp. Res. Board* **2017**, *2645*, 157–167.
- Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F.Y. Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 865–873.
- Polson, N.G.; Sokolov, V.O. Deep learning for short-term traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 1–17.
- Do, L.N.; Vu, H.L.; Vo, B.Q.; Liu, Z.; Phung, D. An effective spatial-temporal attention based neural network for traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2019**, *108*, 12–28.
- Tsekeris, T.; Stathopoulos, A. Real-Time Traffic Volatility Forecasting in Urban Networks. *Transp. Res. Rec. J. Transp. Res. Board* **2006**, *1964*, 146–156.
- Zhang, Y.; Sun, R.; Haghani, A.; Zeng, X. Univariate Volatility-based Models for Improving Quality of Travel Time Reliability Forecasting. *Transp. Res. Rec. J. Transp. Res. Board* **2013**, *2365*, 73–81.

20. Zhang, Y.; Zhang, Y.; Haghani, A. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transp. Res. Part C Emerg. Technol.* **2014**, *43*, 65–78.
21. Hentschel, L. All in the Family Nesting Symmetric and Asymmetric GARCH Models. *J. Financ. Econ.* **1995**, *39*, 71–104.
22. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327.
23. Zakoian, J.M. Threshold heteroskedastic models. *J. Econ. Dyn. Control* **1994**, *18*, 931–955.
24. Higgins, M.L.; Bera, A.K. A class of nonlinear ARCH models. *Int. Econ. Rev.* **1992**, *33*, 137–158.
25. Engle, R.F.; Ng, V.K. Measuring and testing the impact of news on volatility. *J. Financ.* **1993**, *48*, 1749–1778.
26. Engle, R.F. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econom. J. Econom. Soc.* **1982**, *50*, 987–1007.
27. Kamarianakis, Y.; Kanas, A.; Prastacos, P. Modeling Traffic Volatility Dynamics in an Urban Network. *Transp. Res. Rec. J. Transp. Res. Board* **2005**, *1923*, 18–27.
28. Guo, J.; Williams, B.M. Real-Time Short-Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters. *Transp. Res. Rec. J. Transp. Res. Board* **2010**, *2175*, 28–37.
29. Chen, C.; Hu, J.; Meng, Q.; Zhang, Y. Short-time traffic flow prediction with ARIMA-GARCH model. In IEEE Intelligent Vehicles Symposium (IV), IEEE, Baden-Baden, Germany, 5–9 June 2011; pp. 607–612.
30. Ding, C.; Duan, J.; Zhang, Y.; Wu, X.; Yu, G. Using an ARIMA-GARCH modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 1054–1064.
31. Xia, J.; Nie, Q.; Huang, W.; Qian, Z. Reliable Short-term Traffic Flow Forecasting for Urban Roads. *Transp. Res. Rec. J. Transp. Res. Board* **2013**, *2324*, 77–85.
32. Chen, L.; Zheng, L.; Yang, J.; Xia, D.; Liu, W. Short-term traffic flow prediction: From the perspective of traffic flow decomposition. *Neurocomputing* **2020**, *413*, 444–456.
33. Lin, L.; Handley, J.C.; Gu, Y.; Zhu, L.; Wen, X.; Sadek, A.W. Quantifying uncertainty in short-term traffic prediction and its application to optimal staffing plan development. *Transp. Res. Part C Emerg. Technol.* **2018**, *92*, 323–348.
34. Kim, M. Network traffic prediction based on INGARCH model. *Wirel. Netw.* **2020**, *26*, 6189–6202.
35. Yao, R.; Zhang, W.; Zhang, L. Hybrid methods for short-term traffic flow prediction based on ARIMA-GARCH model and wavelet neural network. *J. Transp. Eng. Part A Syst.* **2020**, *146*, 04020086.
36. Lin, X.; Huang, Y. Short-Term High-Speed Traffic Flow Prediction Based on ARIMA-GARCH-M Mode. *Wirel. Pers. Commun.* **2021**, *117*, 3421–3430.
37. Sato, T.; Matsuda, Y. Spatial extension of generalized autoregressive conditional heteroskedasticity models. *Spat. Econ. Anal.* **2021**, *16*, 148–160.
38. Hølleland, S.; Karlsen, H.A. A Stationary Spatio-Temporal GARCH Model. *J. Time Ser. Anal.* **2020**, *41*, 177–209.
39. Zhao, L.; Wen, X.; Wang, Y.; Shao, Y. A novel hybrid model of ARIMA-MCC and CKDE-GARCH for urban short-term traffic flow prediction. *IET Intell. Transp. Syst.* **2022**, *16*, 206–217.
40. Helbing, D.; Schönhof, M.; Kern, D. Volatile Decision Dynamics: Experiments, Stochastic Description, Intermittency Control and Traffic Optimization. *New J. Phys.* **2002**, *4*, 33.
41. Helbing, D.; Batic, D.; Schönhof, M.; Treiber, M. Modelling widely scattered states in ‘synchronized’ traffic flow and possible relevance for stock market dynamics. *Phys. A Stat. Mech. Its Appl.* **2002**, *303*, 251–260.
42. Fuller, W.A. *Introduction to Statistical Time Series*; John Wiley & Sons, Inc.: New York, NY, USA, 2009.
43. Ljung, G.M.; Box, G.E. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297–303.
44. Hyndman, R.J.; Khandakar, Y. Automatic Time Series Forecasting: The forecast Package for R. *J. Stat. Softw.* **2008**, *27*, 1–22.
45. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464.
46. Myung, I.J. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **2003**, *47*, 90–100.
47. Tsay, R.S. *An Introduction to Analysis of Financial Data with R*; John Wiley & Sons, Inc.: New York, NY, USA, 2014.
48. Ding, Z.; Granger, C.W.; Engle, R.F. A Long Memory Property of Stock Market Returns and A New Model. *J. Empir. Financ.* **1993**, *1*, 83–106.
49. Harvey, D.; Leybourne, S.; Newbold, P. Testing the equality of prediction mean squared errors. *Int. J. Forecast.* **1997**, *13*, 281–291.
50. Chen, S.; Härdle, W.K.; Jeong, K. Forecasting volatility with support vector machine-based GARCH model. *J. Forecast.* **2010**, *29*, 406–433.
51. U.S. Department of Transportation. ITS JPO Site: San Diego Test Data Sets. ITS Public Data Hub. 2018. Available online: <https://data.transportation.gov/Automobiles/San-Diego-Test-Data-Sets/6qbu-yvq3> (accessed on 1 August 2022).
52. Pagan, A.R.; Schwert, G.W. Alternative Models for Conditional Shocks Volatility. *J. Econom.* **1990**, *45*, 267–290.